UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/733,178	12/10/2003	Eric P. Jiang	4272.68-13	8293

7590 05/15/2007

Marc A. Hubbard
Munsch Hardt Kopf & Harr, P. C.
4000 Fountain Place
1445 Ross Avenue
Dallas, TX 75202-2790

EXAMINER	
LO, SUZANNE	
ART UNIT	**PAPER NUMBER**
2128	

MAIL DATE	DELIVERY MODE
05/15/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/733,178	JIANG ET AL.	
	Examiner	Art Unit	
	Suzanne Lo	2128	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *12 February 2007*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-56* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-56* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *10 December 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-56 have been presented for examination.

Claim Rejections - 35 USC § 103

2. **Claims 1-56** are rejected under 35 U.S.C. 103(a) as being unpatentable over Applicants' own admission that a method and system automatically performs many or all of the steps of statistical analysis described in the background of the application.

Claims 1-56 appear to be directed to the automation **(page 3 of Specification, [0009])** of a manual activity utilizing the steps disclosed by the Applicant in the background as well as certain sections of the specification of the instant application. *In re Venner*, 262 F.2d 91, 95, 120 USPQ 193, 194 (CCPA 1958), the court held that broadly providing an automatic or mechanical means to replace a manual activity which accomplished the same result is not sufficient to distinguish over the prior art.

As per claim 1, Applicants' own admission is directed to in a computer-based system, a method of building a statistical model, *the method* comprising: automatically identifying and flagging categorical variables in a data set containing both categorical and continuous variables; automatically identifying categorical variables that are correlated with one or more continuous variables and eliminating categorical variables that are correlated with at least one continuous variable from a training data matrix used to build a statistical model, wherein the training data matrix comprises a subset of the original data set; and building the statistical model based on the training data matrix **(page 2 of Specification, [0006])** and therefore known in the art at the time of the invention.

As per claim 2, Applicants' own admission is directed to the method of claim 1 wherein said step of automatically identifying and flagging categorical variables comprises: determining if a variable contains integer observation values; if the variable contains integer values, determining the number of unique integer values contained in the variable; determining if the number of unique values exceeds a

predetermined threshold value; and if the number of unique values does not exceed the threshold value, flagging the variable as a categorical variable **(page 2 of Specification, Sections "Data Exploration" and "Categorical Variable Pre-preprocessing", [0006])** therefore known in the art at the time of the invention.

As per claim 3, Applicants' own admission is directed to the method of claim 2 further comprising: if the number of unique values exceeds the threshold value, determining if the variable has predictive strength greater than a predetermined value of Pearson's r; if the variable has predictive strength greater than the predetermined value of Pearson's r, flagging the variable as a continuous variable; if the variable has predictive strength less than the predetermined value of Pearson's r, reducing the number of unique values by eliminating those unique values containing less than a predetermined number of entries so as to create a reduced variable set with a reduced number of unique values; determining if the reduced number of unique values exceeds the threshold value; and if the reduced number of unique values does not exceed the threshold value, flagging the variable as a categorical variable, else flagging the variable as a continuous variable **(page 10, [0040] and page 2 of Specification, Sections "Data Exploration" and "Categorical Variable Pre-preprocessing", [0006])** therefore known in the art at the time of the invention.

As per claim 4, Applicants' own admission is directed to the method of claim 1 wherein said step of automatically identifying categorical variables that are correlated with one or more continuous variables comprises: binning at least one continuous variable so as to convert the continuous variable into a psuedo-categorical variable; and calculating a Cramer's V value between at least one categorical variable and the psuedo-categorical variable to obtain an estimated measure of co-linearity between the categorical variable and the continuous variable **(page 10, [0040] and page 2 of Specification, Sections "Data Exploration" and "Categorical Variable Pre-preprocessing", [0006])** therefore known in the art at the time of the invention.

As per claim 5, Applicants' own admission is directed to the method of claim 1 further comprising: calculating a correlation value for each variable in the training data matrix with respect to a target variable; sorting the variables based on their correlation with the target variable; and retaining a predetermined number of variables having the highest correlation values and eliminating any remaining variables from the training data matrix **(page 2 of Specification, Section "Variable Reduction" [0006] and page 10, [0042])** therefore known in the art at the time of the invention.

As per claim 6, Applicants' own admission is directed to the method of claim 1 further comprising: expanding each categorical variable contained in the training data matrix into a plurality of dummy variables; measuring a predictive strength for each dummy variable and continuous variable in the training data matrix toward a target variable; determining if any pair of variables in the set of dummy and continuous variables exhibits a pair-wise correlation greater than a predetermined threshold; and if a pair of variables exhibits a pair-wise correlation greater than the threshold, eliminating one of the variables in the pair from the training data matrix, wherein the eliminated variable exhibits less predictive strength toward the target variable than the non-eliminated variable in the pair **(page 2 of Specification, Section "Variable Reduction" [0006] and page 10, [0042])** therefore known in the art at the time of the invention.

As per claim 7, Applicants' own admission is directed to the method of claim 1 further comprising: creating a plurality of principle components from the variables contained in the training data matrix, wherein each principle component comprises a linear combination of variables; sorting the plurality of principle components by how much variance of the training data matrix each component captures; selecting a subset of the plurality of principle components that captures a variance greater than a predetermined percentage of total variance; and using the selected principle components to build the statistical model **(page 2 of Specification, [0006], Sections "Create Model" and "Model Selection" and page 18, [0069]-[0072], and [0076])** therefore known in the art at the time of the invention.

As per claim 8, Applicants' own admission is directed to the method of claim 7 wherein said step of using the selected principle components to build the statistical model comprises: performing a singular value decomposition (SVD) to generate a loading matrix; and mapping coefficients calculated for the principle components back to corresponding variables of the training data matrix using the loading matrix **(page 2 of Specification, [0006], Sections "Create Model" and "Model Selection" and page 18, [0069]-[0072], and [0076])** therefore known in the art at the time of the invention.

As per claim 9, Applicants' own admission is directed to the method of claim 1 further comprising: performing a singular value decomposition (SVD) analysis using the variables contained in the training data matrix if the number of records in the training data matrix is less than a predetermined value; and otherwise, performing a conjugate gradient descent (CGD) analysis on a residual sum of squares based on the variables contained in the training data matrix if the number of records in the training data matrix is greater than or equal to the predetermined value **(page 2 of Specification, [0006], Sections "Create Model" and "Model Selection" and page 18, [0069]-[0072], and [0076])** therefore known in the art at the time of the invention.

As per claim 10, Applicants' own admission is directed to the method of claim 1 further comprising: detecting outlier values in the data set; and for each detected outlier value, presenting a user with the following three options for handling the outlier value: (1) substitute the outlier value with a maximum or minimum non-outlier value in the data set; (2) keep the outlier value in the data set; (3) delete the record corresponding to the outlier value **(page 2 of Specification, [0006], Section "Data Cleansing", and page 11, [0044])** therefore known in the art at the time of the invention.

As per claim 11, Applicants' own admission is directed to the method of claim 1 further comprising: detecting missing values in the data set; and for each missing value of a variable, inserting amean value of non-missing values of the variable in place of the missing value in the data set **(page 2 of Specification, [0006], Section "Data Cleansing")** therefore known in the art at the time of the invention.

As per claim 12, Applicants' own admission is directed to the method of claim 1 further comprising: automatically detecting continuous variables having an exponential distribution; and log-scaling those continuous variables using the following formula:x(i)—min bx(i) = 1 – e mean—min where x(i) is a continuous variable being analyzed, min, and mean is the minimum value and the mean value of the variable in samples, respectively **(page 2 of Specification, Section "Variable Standardization and page 13, [0057]-[0053])** therefore known in the art at the time of the invention.

As per claim 13, Applicants' own admission is directed to the method of claim 12 further comprising normalizing all the variables in the training data matrix **(page 2 of Specification, Section "Variable Standardization)** therefore known in the art at the time of the invention.

As per claim 14, Applicants' own admission is directed to the method of claim 1 further comprising randomly splitting the data set into a subset of training variables and a subset of test variables, wherein the training variables are used to create the training data matrix for building the model and the subset of test variables are subsequently used to test the resulting model **(page 2 of Specification, Sections "Split Data Set" and "Model Validation")** therefore known in the art at the time of the invention.

As per claim 15, Applicants' own admission is directed to the method of claim 14 wherein prior to using the subset of test variables to test the model, pre-processing is performed on variables in the test set so as to create a test data matrix containing the same variables and same format as the training data matrix **(page 2 of Specification, [0006], Sections "Data Exploration" and "Categorical Variable Preprocessing")** therefore known in the art at the time of the invention.

As per claims 16-28, the claims are directed to methods with the same limitations as claims 1-15 and therefore rejected over the same art.

As per claims 29-56, the claims are directed to a computer-readable medium containing code when executed performs method steps with the same limitations as claims 1-28 and therefore rejected over the same art.

3. **Claims 1-3, 6, 12-13, 16-17, 21-23, 28-31, 33, 40-41, 44-45, 50-51, and 56** are rejected under 35 U.S.C. 103(a) as being unpatentable over **Wang et al. (U.S. Patent No. 6,470,229 B1) in view of Brown et al. (U.S. Patent No. 6,473,080 B1)**.

As per claim 1, Wang is directed to in a computer-based system **(column 12, line 66 – column 3, line 1)**, a method of building a statistical model, *the method* comprising: automatically identifying and flagging categorical variables in a data set containing both categorical and continuous variables **(column 3, lines 60-67)**; automatically identifying categorical variables that are correlated with one or more continuous variables and eliminating categorical variables that are correlated with at least one continuous variable from training data used to build a statistical model **(column 4, lines 27-35)**, wherein the training data comprises a subset of the original data set **(column 4, lines 36-42)**; and building the statistical model based on the training data **(column 3, lines 49-53)** but fails to specifically disclose the training data in a matrix. Brown teaches organizing data in a matrix **(column 6, lines 41-53)**. Wang and Brown are analogous art because they are from the same field of endeavor building statistical models. It would have been obvious to an ordinary person skilled in the art at the time of the invention to combine the statistical model building method with the data organization of Brown in order to creating a data architecture that is easily navigable **(Brown, column 5, lines 20-21)**.

As per claim 2, the combination of Wang and Brown already discloses the method of claim 1 wherein said step of automatically identifying and flagging categorical variables comprises: determining if a variable contains integer observation values; if the variable contains integer values, determining the number of unique integer values contained in the variable; determining if the number of unique values

exceeds a predetermined threshold value; and if the number of unique values does not exceed the threshold value, flagging the variable as a categorical variable **(Wang, column 4, lines 8-35)**.

As per claim 3, the combination of Wang and Brown already discloses the method of claim 2 further comprising: if the number of unique values exceeds the threshold value, determining if the variable has predictive strength greater than a predetermined value of Pearson's r; if the variable has predictive strength greater than the predetermined value of Pearson's r, flagging the variable as a continuous variable; if the variable has predictive strength less than the predetermined value of Pearson's r, reducing the number of unique values by eliminating those unique values containing less than a predetermined number of entries so as to create a reduced variable set with a reduced number of unique values; determining if the reduced number of unique values exceeds the threshold value; and if the reduced number of unique values does not exceed the threshold value, flagging the variable as a categorical variable, else flagging the variable as a continuous variable **(Brown, column 12, lines 15-35)**.

As per claim 5, the combination of Wang and Brown already discloses the method of claim 1 further comprising: expanding each categorical variable contained in the training data matrix into a plurality of dummy variables; measuring a predictive strength for each dummy variable and continuous variable in the training data matrix toward a target variable; determining if any pair of variables in the set of dummy and continuous variables exhibits a pair-wise correlation greater than a predetermined threshold; and if a pair of variables exhibits a pair-wise correlation greater than the threshold, eliminating one of the variables in the pair from the training data matrix, wherein the eliminated variable exhibits less predictive strength toward the target variable than the non-eliminated variable in the pair **(Wang, column 10, lines 32-63)**.

As per claim 12, the combination of Wang and Brown already discloses the method of claim 1 further comprising: automatically detecting continuous variables having an exponential distribution; and log-scaling those continuous variables using the following formula:x(i)—min bx(i) = 1 – e mean—min

where x(i) is a continuous variable being analyzed, min, and mean is the minimum value and the mean value of the variable in samples, respectively **(Brown, column 9, lines 22-33 and column 10, lines 36-45)**.

As per claim 13, the combination of Wang and Brown already discloses the method of claim 12 further comprising normalizing all the variables in the training data matrix **(Brown, column 9, lines 22-33)**.

As per claims 16-17, 21-22, 23, and 28, the claims are directed to methods with the same limitations as claims 1-3, 6, and 12-13 above and therefore rejected under the same art combination.

As per claims 29-31, 33, 40-41, 44-45, 50-51, and 56, the claims are directed to a computer-readable medium containing code when executed performs method steps with the same limitations as claims 1-3, 6, and 12-13 above and therefore rejected under the same art combination.

4. **Claims 7-8, 18-19, 25-26, 35-36, 46-47, 49, and 53-54** are rejected under 35 U.S.C. 103(a) as being unpatentable over Wang et al. (U.S. Patent No. 6,470,229 B1) and Brown et al. (U.S. Patent No. 6,473,080 B1) **in further view of Vaithyanathan et al (U.S. Patent No. 5,819,258)**.

As per claim 7, the combination of Brown and Wang is directed to the method of claim 1 but fails to specifically disclose further comprising: creating a plurality of principle components from the variables contained in the training data matrix, wherein each principle component comprises a linear combination of variables; sorting the plurality of principle components by how much variance of the training data matrix each component captures; selecting a subset of the plurality of principle components that captures a variance greater than a predetermined percentage of total variance; and using the selected principle components to build the statistical model. Vaithyanathan teaches using the method of principle component analysis **(column 8, line 61-column 9, line 4)**. Brown, Wang, and Vithyanathan are analogous art because they are all from the same field of endeavor, building a statistical model. It would

have been obvious to an ordinary person skilled in the art at the time of the invention to combine the statistical model building method of Wang and Brown with the PCA method of Vaithyanathan in order to reduce the data set for manageability **(Vaithyanathan, column 8, lines 61-67)**.

As per claim 8, the combination of Wang, Brown, and Vithyanathan already discloses the method of claim 7 wherein said step of using the selected principle components to build the statistical model comprises: performing a singular value decomposition (SVD) to generate a loading matrix; and mapping coefficients calculated for the principle components back to corresponding variables of the training data matrix using the loading matrix **(column 9, lines 5-65)**.

As per claims 18-19 and 25-26, the claims are directed to methods with the same limitations as claims 7-8 above and therefore rejected under the same art combination.

As per claims 35-36, 46-47, 49, and 53-54, the claims are directed to a computer-readable medium containing code when executed performs method steps with the same limitations as claims 7-8 and 12 above and therefore rejected under the same art combination.

Response to Arguments

5. Objection to the oath has been withdrawn; claim objections as well as 101 and 112 rejections have been withdrawn due to the amended claims.

6. Applicant's arguments filed 02/12/07 have been fully considered but they are not persuasive.

7. In response to Applicant's argument that Applicant's own admission does not teach "automatically identifying categorical variables that are correlated with one or more continuous variables and eliminating categorical variables that are correlated with at least one continuous variable", Applicant is further directed to paragraphs **[0040], [0042], and [0050]** of the Specification of the instant application which discloses identifying categorical variables and the corresponding correlations of said variables. Applicant's are also directed to paragraphs **[0006]-[0009]** of the Specification of the instant application

which after identifying the correlating variables in Categorical Variable Preprocessing, Variable Reduction and Model Creation involves "deciding which variables should be included in created a statistical model for a given target variable and which variables should be excluded" and "[making] decisions as to whether the data is continuous, categorical, highly predictive, or redundant" which are performed automatically.

Furthermore, Applicant's own admission, regardless of where in the specification it is located, is understood as known in the art and given up for public use. While the following passages are located in the section titled Detailed Description of the Preferred Embodiments, the language appears to disavow novelty of the disclosed method steps.

In paragraph **[0029]**: "*As well known in the art,* the data set 10, or at least a subset thereof, can be used as "training data" to create a statistical model that provides a predictive correlation between the predictive variables and the target variable".

In paragraph **[0033]**: "*In the art of statistical analysis,* two common types of variables are "categorical" and continuous" variables. The characteristics and differences between these two types of variables are *well known in the art.*"

In paragraph **[0050]**: "All of these statistical measures are *standard and well-known* formulas may be used to calculate their values."

In paragraph **[0051]**: "During the exploratory data analysis phase of a modeling project, *statisticians frequently* encounter variables that might reasonably be assumed to have an exponential distribution (e.g. monthly household income). *Statisticians will often handle* this situation by transforming the variable to a logarithmic scale to prior to model building."

8. Wang teaches identifying correlated variables, **column 10, lines 31-62** and choosing only one variable which is more relevant to derive a split in the logic tree, the other variable is disregarded, even

with a categorical variable correlated with a continuous variable, the categorical variable is inherently eliminated **column 7, lines 24-54**.

Conclusion

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. These references include:

1. U.S. Patent No. 5,781,430 issued Tsai on 07/14/98.

2. U.S. Patent No. 5,452,410 issued Magidson on 09/19/98.

10. All Claims are rejected.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Suzanne Lo whose telephone number is (571)272-5876. The examiner can normally be reached on M-F, 8-4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kamini Shah can be reached on (571)272-2297. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Suzanne Lo
Patent Examiner
Art Unit 2128

SL
05/07/07

KAMINI SHAH
SUPERVISORY PATENT EXAMINER